如何在CDH集群启用Kerberos

原创： Fayson [Hadoop实操](javascript:void(0);) 2017-09-16

温馨提示：要看高清无码套图，请使用手机打开并单击图片放大查看。

**1.文档编写目的**

本文档讲述如何在CDH集群启用及配置Kerberos，您将学习到以下知识：

1.如何安装及配置KDC服务

2.如何通过CDH启用Kerberos

3.如何登录Kerberos并访问Hadoop相关服务

文档主要分为以下几步：

1.安装及配置KDC服务

2.CDH集群启用Kerberos

3.Kerberos使用

这篇文档将重点介绍如何在CDH集群启用及配置Kerberos，并基于以下假设：

1.CDH集群运行正常

2.集群未启用Kerberos

3.MySQL 5.1.73

以下是本次测试环境，但不是本操作手册的必需环境：

1.操作系统：CentOS 6.5

2.CDH和CM版本为5.12.0

3.采用root用户进行操作

**2.KDC服务安装及配置**

本文档中将KDC服务安装在Cloudera Manager Server所在服务器上（KDC服务可根据自己需要安装在其他服务器）

1.在Cloudera Manager服务器上安装KDC服务

*[root@ip-172-31-6-148*~]# yum -y install krb5-serverkrb5-libs krb5-auth-dialog krb5-workstation

--C7有差别

2.修改/etc/krb5.conf配置

*[root@ip-172-31-6-148* fayson\_r]# vim /etc/krb5.conf   
*[logging]  
 default* = FILE:/var/log/krb5libs.log  
 *kdc* = FILE:/var/log/krb5kdc.log  
 *admin\_server* = FILE:/var/log/kadmind.log  
  
*[libdefaults]  
 default\_realm* = FAYSON.COM  
 *dns\_lookup\_realm* = false  
 *dns\_lookup\_kdc* = false  
 *ticket\_lifetime* = 24h  
 *renew\_lifetime* = 7d  
 *forwardable* = true  
  
*[realms]  
 FAYSON.COM* = {  
  *kdc* = ip-172-31-6-148.fayson.com  
  *admin\_server* = ip-172-31-6-148.fayson.com  
 }  
  
*[domain\_realm]  
 .ip-172-31-6-148.fayson.com* = FAYSON.COM  
 *ip-172-31-6-148.fayson.com* = FAYSON.COM

标红部分为需要修改的信息。

3.修改/var/kerberos/krb5kdc/kadm5.acl配置

*[root@ip-172-31-6-148*~]# vim /var/kerberos/krb5kdc/kadm5.acl   
*\*/admin@FAYSON.COM*\*

4.修改/var/kerberos/krb5kdc/kdc.conf配置

*[root@ip-172-31-6-148* ~]# vim /var/kerberos/krb5kdc/kdc.conf   
*[kdcdefaults]  
 kdc\_ports* = 88  
 *kdc\_tcp\_ports* = 88  
  
*[realms]  
 FAYSON.COM*= {  
  #master\_key\_type = aes256-cts  
  max\_renewable\_life= *7d* 0h 0m 0s  
  *acl\_file* = /var/kerberos/krb5kdc/kadm5.acl  
  *dict\_file* = /usr/share/dict/words  
  *admin\_keytab* = /var/kerberos/krb5kdc/kadm5.keytab  
  *supported\_enctypes* = aes256-cts:normal aes128-cts:normal des3-hmac-sha1:normal arcfour-hmac:normal des-hmac-sha1:n  
*ormal* des-cbc-md5:normal des-cbc-crc:normal  
 }

标红部分为需要修改的配置。

5.创建Kerberos数据库

*[root@ip-172-31-6-148* ~]# kdb5\_util create –r *FAYSON.COM* -s  
*Loading* random data  
*Initializing* database '/var/kerberos/krb5kdc/principal' **for** realm 'FAYSON.COM',  
*master* key name 'K/M@FAYSON.COM'  
*You* will be prompted **for** the database Master Password.  
*It* is important that you NOT FORGET this password.  
*Enter* KDC database master key:   
*Re-enter* KDC database master key to verify:

此处需要输入Kerberos数据库的密码。

6.创建Kerberos的管理账号

*[root@ip-172-31-6-148* ~]# kadmin.local  
*Authenticating* as principal fayson/admin@CLOUDERA.COM with password.  
*kadmin.local:* addprinc admin/admin@FAYSON.COM  
*WARNING:* no policy specified **for** admin/admin@FAYSON.COM; *defaulting* to no policy  
*Enter* password **for** principal "admin/admin@FAYSON.COM":   
*Re-enter* password **for** principal "admin/admin@FAYSON.COM":   
*Principal* "admin/admin@FAYSON.COM" created.  
*kadmin.local:* exit  
*[root@ip-172-31-6-148* ~]#

标红部分为Kerberos管理员账号，需要输入管理员密码。

7.将Kerberos服务添加到自启动服务，并启动krb5kdc和kadmin服务

*[root@ip-172-31-6-148*~]# chkconfig krb5kdc on  
*[root@ip-172-31-6-148*~]# chkconfig kadmin on  
*[root@ip-172-31-6-148*~]# service krb5kdc start  
*Starting*Kerberos 5 KDC:                                   **[**OK **]***[root@ip-172-31-6-148*~]# service kadmin start  
*Starting*Kerberos 5 Admin Server:                          **[**OK **]***[root@ip-172-31-6-148*~]#

8.测试Kerberos的管理员账号

*[root@ip-172-31-6-148* ~]# kinit admin/admin@FAYSON.COM  
*Password* **for** admin/admin@FAYSON.COM:   
*[root@ip-172-31-6-148* ~]# klist  
*Ticket* cache: FILE:/tmp/krb5cc\_0  
*Default* principal: admin/admin@FAYSON.COM  
  
*Valid* starting     Expires            Service principal  
*09/05/17* 16:39:17  09/06/17 16:39:17  krbtgt/FAYSON.COM@FAYSON.COM  
        *renew* **until** 09/12/17 16:39:17  
*[root@ip-172-31-6-148* ~]#

9.为集群安装所有Kerberos客户端，包括Cloudera Manager

*[root@ip-172-31-6-148* cdh-shell-master]# yum -y install krb5-libs krb5-workstation

10.在Cloudera Manager Server服务器上安装额外的包

*[root@ip-172-31-6-148*cdh-shell-master]# yum -y install openldap-clients

11.将KDC Server上的krb5.conf文件拷贝到所有Kerberos客户端

*[root@ip-172-31-6-148*cdh-shell-master]# scp -r /etc/krb5.conf root@172.31.5.190:/etc/

此处使用脚本进行拷贝

*[root@ip-172-31-6-148*cdh-shell-master]# sh b.sh node.list /etc/krb5.conf  /etc/  
*krb5.conf*100% 451    0.4KB/s   00:00     
*krb5.conf*100% 451    0.4KB/s   00:00     
*krb5.conf*100% 451    0.4KB/s   00:00     
*krb5.conf*100%  451    0.4KB/s  00:00      
*[root@ip-172-31-6-148*cdh-shell-master]#

**3.CDH集群启用Kerberos**

1.在KDC中给Cloudera Manager添加管理员账号

*[root@ip-172-31-6-148* cdh-shell-bak]# kadmin.local  
*Authenticating* as principal admin/admin@FAYSON.COM with password.  
*kadmin.local:* addprinc cloudera-scm/admin@FAYSON.COM  
*WARNING:* no policy specified **for** cloudera-scm/admin@FAYSON.COM; *defaulting* to no policy  
*Enter* password **for** principal "cloudera-scm/admin@FAYSON.COM":   
*Re-enter* password **for** principal "cloudera-scm/admin@FAYSON.COM":   
*Principal* "cloudera-scm/admin@FAYSON.COM" created.  
*kadmin.local:* exit  
*[root@ip-172-31-6-148* cdh-shell-bak]#

2.进入Cloudera Manager的“管理”-> “安全”界面

3.选择“启用Kerberos”，进入如下界面

确保如下列出的所有检查项都已完成

4.点击“继续”，配置相关的KDC信息，包括类型、KDC服务器、KDC Realm、加密类型以及待创建的Service Principal（hdfs，yarn,，hbase，hive等）的更新生命期等

5.点击“继续”

6.不建议让Cloudera Manager来管理krb5.conf, 点击“继续”

7.输入Cloudera Manager的Kerbers管理员账号，必须和之前创建的账号一致，点击“继续”

8.等待启用Kerberos完成，点击“继续”

9.点击“继续”

10.勾选重启集群，点击“继续”

11.等待集群重启成功，点击“继续”

至此已成功启用Kerberos。

**4.Kerberos使用**

使用fayson用户运行MapReduce任务及操作Hive，需要在集群所有节点创建fayson用户。

1.使用kadmin创建一个fayson的principal

*[root@ip-172-31-6-148* cdh-shell-bak]# kadmin.local  
*Authenticating* as principal admin/admin@FAYSON.COM with password.  
*kadmin.local:* addprinc fayson@FAYSON.COM  
*WARNING:* no policy specified **for** fayson@FAYSON.COM; *defaulting* to no policy  
*Enter* password **for** principal "fayson@FAYSON.COM":  
*Re-enter* password **for** principal "fayson@FAYSON.COM":  
*Principal* "fayson@FAYSON.COM" created.  
*kadmin.local:* exit  
*[root@ip-172-31-6-148* cdh-shell-bak]#

2.使用fayson用户登录Kerberos

*[root@ip-172-31-6-148* cdh-shell-bak]# kdestroy  
*[root@ip-172-31-6-148* cdh-shell-bak]# kinit fayson  
*Password* **for** fayson@FAYSON.COM:   
*[root@ip-172-31-6-148* cdh-shell-bak]# klist  
*Ticket* cache: FILE:/tmp/krb5cc\_0  
*Default* principal: fayson@FAYSON.COM  
  
*Valid* starting     Expires            Service principal  
*09/05/17* 17:19:08  09/06/17 17:19:08  krbtgt/FAYSON.COM@FAYSON.COM  
        *renew* **until** 09/12/17 17:19:08  
*[root@ip-172-31-6-148* cdh-shell-bak]#

3.运行MapReduce作业

*[root@ip-172-31-6-148*~]# hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-0.20-mapreduce/hadoop-examples.jar pi 10 1  
*...  
Starting*Job  
*17/09/02*20:10:43 INFO mapreduce.Job: Running job: job\_1504383005209\_0001  
*17/09/02*20:10:56 INFO mapreduce.Job: Job job\_1504383005209\_0001 running **in**ubermode : false  
*17/09/02*20:10:56 INFO mapreduce.Job:  map0% reduce 0%  
*17/09/02*20:11:09 INFO mapreduce.Job:  map20% reduce 0%  
*17/09/02*20:11:12 INFO mapreduce.Job:  map40% reduce 0%  
*17/09/02*20:11:13 INFO mapreduce.Job:  map50% reduce 0%  
*17/09/02*20:11:15 INFO mapreduce.Job:  map60% reduce 0%  
*17/09/02*20:11:16 INFO mapreduce.Job:  map70% reduce 0%  
*17/09/02*20:11:19 INFO mapreduce.Job:  map80% reduce 0%  
*17/09/02*20:11:21 INFO mapreduce.Job:  map100% reduce 0%  
*17/09/02*20:11:26 INFO mapreduce.Job:  map100% reduce 100%  
*17/09/02*20:11:26 INFO mapreduce.Job: Job job\_1504383005209\_0001 completedsuccessfully

4.使用beeline连接hive进行测试

*[root@ip-172-31-6-148* cdh-shell-bak]# beeline   
*Beeline* version 1.1.0-cdh5.12.1 by Apache Hive  
*beeline> !connect* jdbc:hive2://localhost:10000/;principal=hive/ip-172-31-6-148.fayson.com@FAYSON.COM  
*...  
Transaction* isolation: TRANSACTION\_REPEATABLE\_READ  
*0:* jdbc:hive2://localhost:10000/*> show* tables;  
*...  
INFO* : OK  
*+-------------+--+*|  *tab\_name* |  
*+-------------+--+*| *test\_table* |  
*+-------------+--+  
1* row selected (0.194 *seconds*)  
0: jdbc:hive2://localhost:10000/*>* **select** \* from test\_table;  
*...  
INFO* : OK  
*+----------------+----------------+--+*| *test\_table.s1* | *test\_table.s2* |  
*+----------------+----------------+--+*| *4* | *lisi* |  
| *1* | *test* |  
| *2* | *fayson* |  
| *3* | *zhangsan* |  
*+----------------+----------------+--+  
4* rows selected (0.144 *seconds*)  
0: jdbc:hive2://localhost:10000/*>*

运行Hive MapReduce作业

0: jdbc:hive2://localhost:10000/*>* **select** count(\*) *from* test\_table;  
*...  
INFO* : OK  
*+------+--+*| *\_c0* |  
*+------+--+*| *4* |  
*+------+--+  
1* row selected (35.779 *seconds*)  
0: jdbc:hive2://localhost:10000/*>*

**5.常见问题**

1.使用Kerberos用户身份运行MapReduce作业报错

*main*: run as user is fayson  
*main*: requested yarn user is fayson  
*Requested*user fayson is not whitelisted and has id 501,whichis below the minimum allowed 1000  
  
*Failing*this attempt. Failing the application.  
*17/09/02*20:05:04 INFO mapreduce.Job: Counters: 0  
*Job*Finished **in**6.184 seconds  
*java.io.FileNotFoundException:*File does not exist:hdfs://ip-172-31-6-148:8020/user/fayson/QuasiMonteCarlo\_1504382696029\_1308422444/out/reduce-out  
        *at*org.apache.hadoop.hdfs.DistributedFileSystem$20.doCall(DistributedFileSystem.java:1266)  
        *at*org.apache.hadoop.hdfs.DistributedFileSystem$20.doCall(DistributedFileSystem.java:1258)  
        *at*org.apache.hadoop.fs.FileSystemLinkResolver.resolve(FileSystemLinkResolver.java:81)  
        *at*org.apache.hadoop.hdfs.DistributedFileSystem.getFileStatus(DistributedFileSystem.java:1258)  
        *at*org.apache.hadoop.io.SequenceFile$Reader.*<init>(SequenceFile.java:1820)  
        at*org.apache.hadoop.io.SequenceFile$Reader.*<init>(SequenceFile.java:1844)  
        at*org.apache.hadoop.examples.QuasiMonteCarlo.estimatePi(QuasiMonteCarlo.java:314)  
        *at*org.apache.hadoop.examples.QuasiMonteCarlo.run(QuasiMonteCarlo.java:354)  
        *at*org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)  
        *at*org.apache.hadoop.examples.QuasiMonteCarlo.main(QuasiMonteCarlo.java:363)  
        *at*sun.reflect.NativeMethodAccessorImpl.invoke0(Native*Method*)  
        at *sun.reflect.NativeMethodAccessorImpl.invoke*(NativeMethodAccessorImpl.java:57)  
        *at*sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)  
        *at*java.lang.reflect.Method.invoke(Method.java:606)  
        *at*org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)  
        *at*org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)  
        *at*org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)  
        *at*sun.reflect.NativeMethodAccessorImpl.invoke0(Native*Method*)  
        at *sun.reflect.NativeMethodAccessorImpl.invoke*(NativeMethodAccessorImpl.java:57)  
        *at*sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)  
        *at*java.lang.reflect.Method.invoke(Method.java:606)  
        *at*org.apache.hadoop.util.RunJar.run(RunJar.java:221)  
        *at*org.apache.hadoop.util.RunJar.main(RunJar.java:136)

问题原因：是由于Yarn限制了用户id小于10000的用户提交作业；

解决方法：修改Yarn的min.user.id来解决